Exploratory Data Analysis

Abril Gonzalez

2024-09-26

```
depress <- read.table("/Users/abrilgonzalez/Desktop/math 130/data/Depress.txt", header = TRUE, sep = "\t
```

```
library(ggplot2)
```

Introduction

What variables are going to be explored?

The data being analyzed in this exploratory data analysis are the factors that can be potentially correlated to depression. This data set has 294 observations from various individuals and contains demographical and socioeconomic data. In this analysis though I will be focusing specifically on the gender, marital status, employment status, and depression levels by looking at the cases variable. This data set was one of the options provided in the guidelines document for this project.

Research Question

What patterns or trends can be observed between gender, marital status, and employment status in different cases of depression?

Hypothesis:

1. Women will report higher levels of depression compared to men.
2. Single, divorced, or widowed individuals will report higher levels of depression compared to married individuals.
3. Employed individuals will report lower levels of depression compared to unemployed individuals.

Univariate Description

Categorical Values

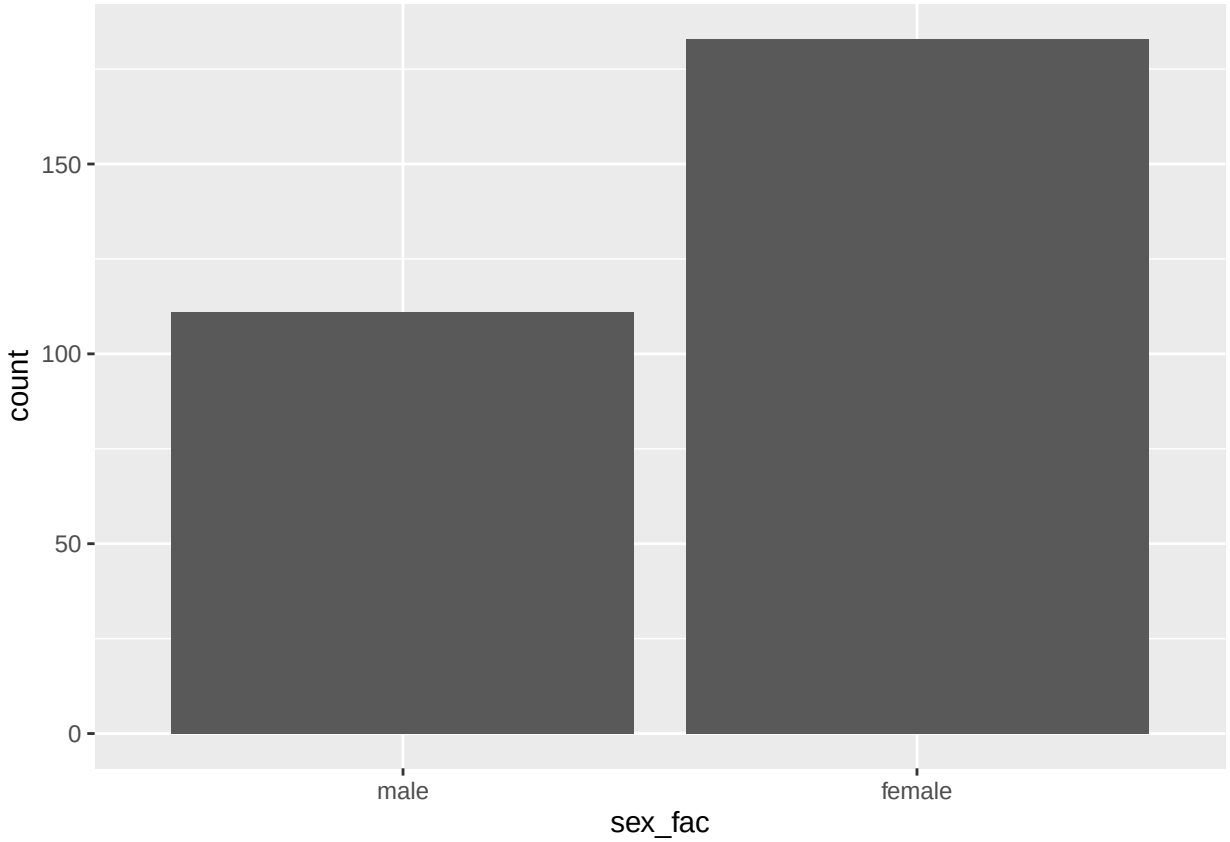
Gender

```
depress$sex_fac <- factor(depress$SEX, labels=c("male", "female"))
```

```
table(depress$SEX, depress$sex_fac)
```

```
##
##     male female
##   1  111      0
##   2    0    183
```

```
ggplot(depress, aes(x=sex_fac)) + geom_bar()
```



There is a total of 111 males and 183 females.

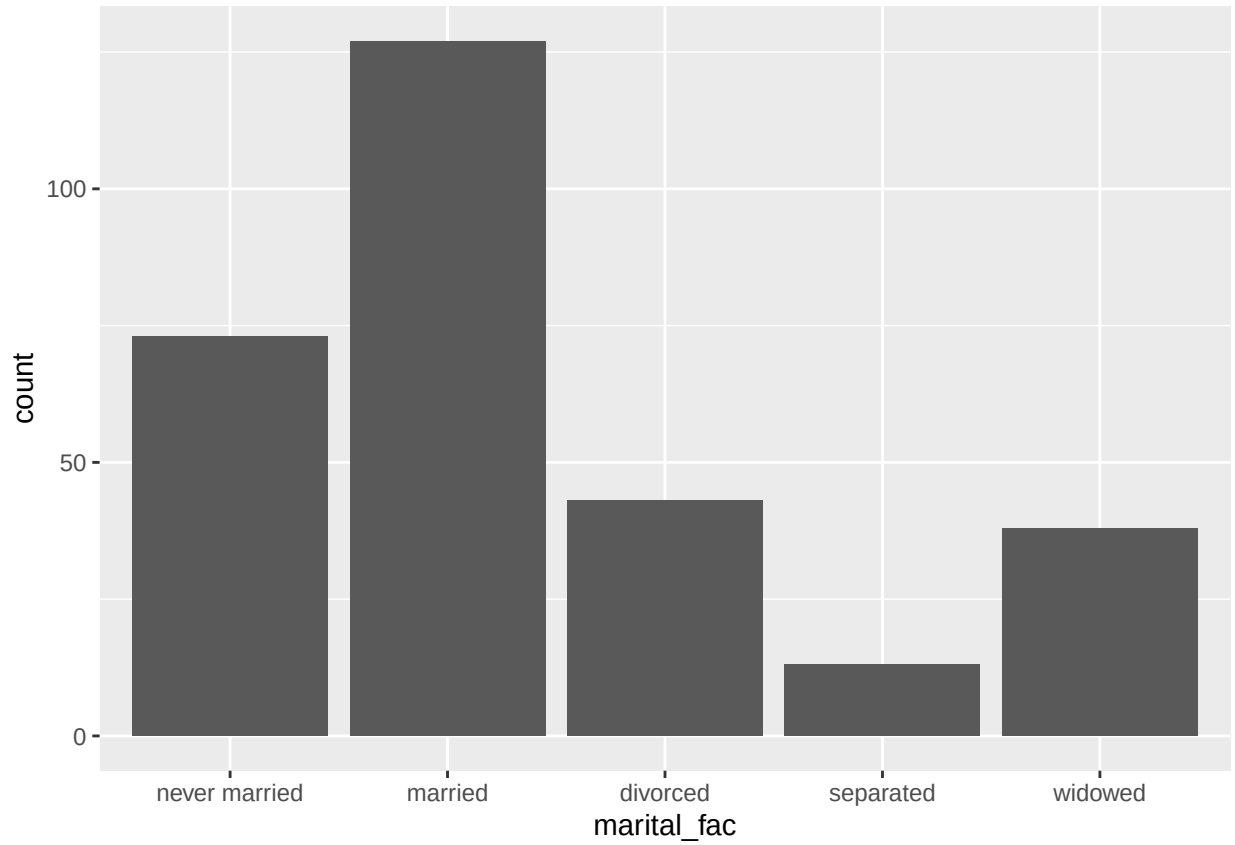
Marital Status

```
depress$marital_fac <- factor(depress$MARITAL, labels=c("never married", "married", "divorced", "separa
```

```
table(depress$MARITAL, depress$marital_fac)
```

```
##
##     never married married divorced separated widowed
##   1            73       0        0         0       0
##   2             0     127        0         0       0
##   3             0       0       43         0       0
##   4             0       0        0        13       0
##   5             0       0        0         0      38
```

```
ggplot(depress, aes(x=marital_fac)) + geom_bar()
```



The majority of the individuals are married. 73 have never been married, 127 are married, 43 are divorced, 13 are separated, and 38 are widowed.

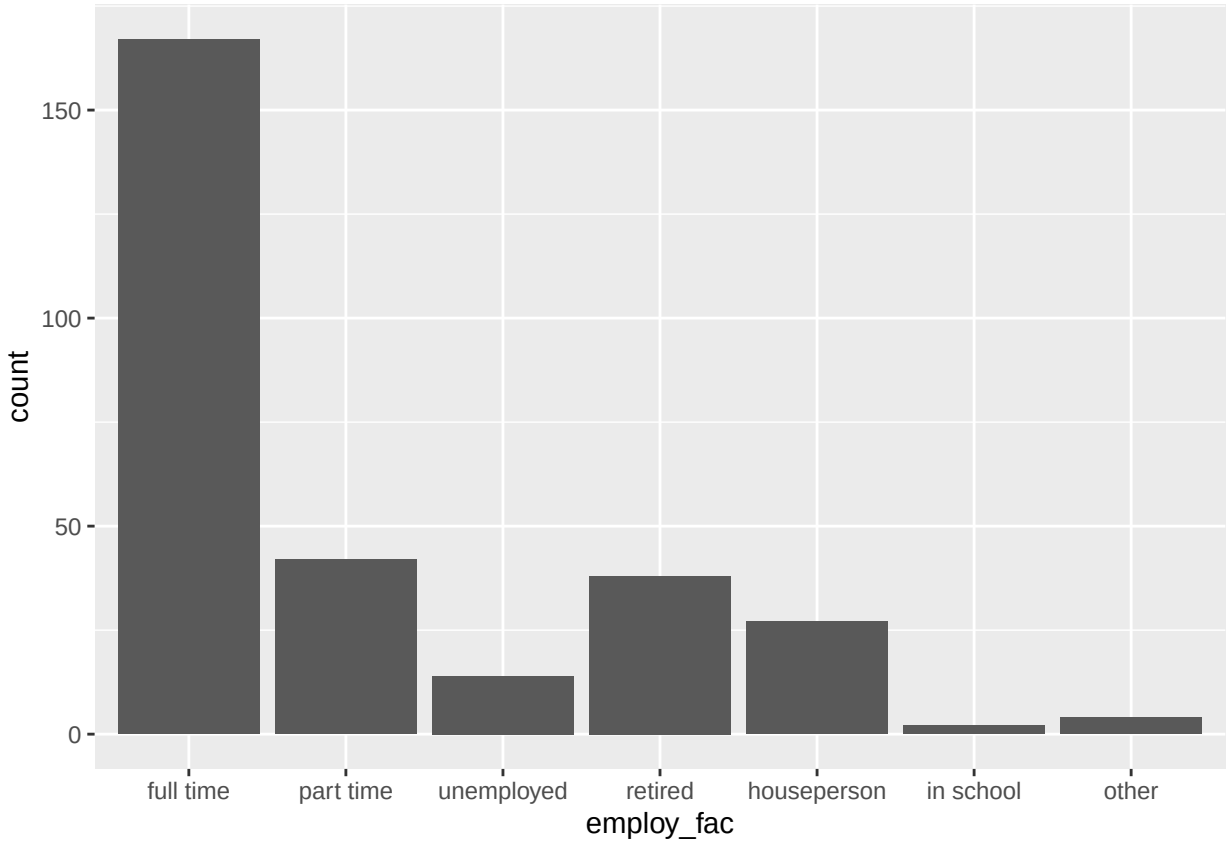
Employment Status

```
depress$employ_fac <- factor(depress$EMPLOY, labels=c("full time", "part time", "unemployed", "retired"
```

```
table(depress$EMPLOY, depress$employ_fac)
```

```
##
##     full time part time unemployed retired houseperson in school other
##   1       167         0          0       0           0         0     0
##   2         0        42          0       0           0         0     0
##   3         0         0         14       0           0         0     0
##   4         0         0          0      38           0         0     0
##   5         0         0          0       0          27         0     0
##   6         0         0          0       0           0         2     0
##   7         0         0          0       0           0         0     4
```

```
ggplot(depress, aes(x=employ_fac)) + geom_bar()
```



The majority of individuals are employed full time. 167 are full-time employees, 42 are part-time, 14 are unemployed, 38 are retired, 27 are housepeople, 2 are in school, 4 are other.

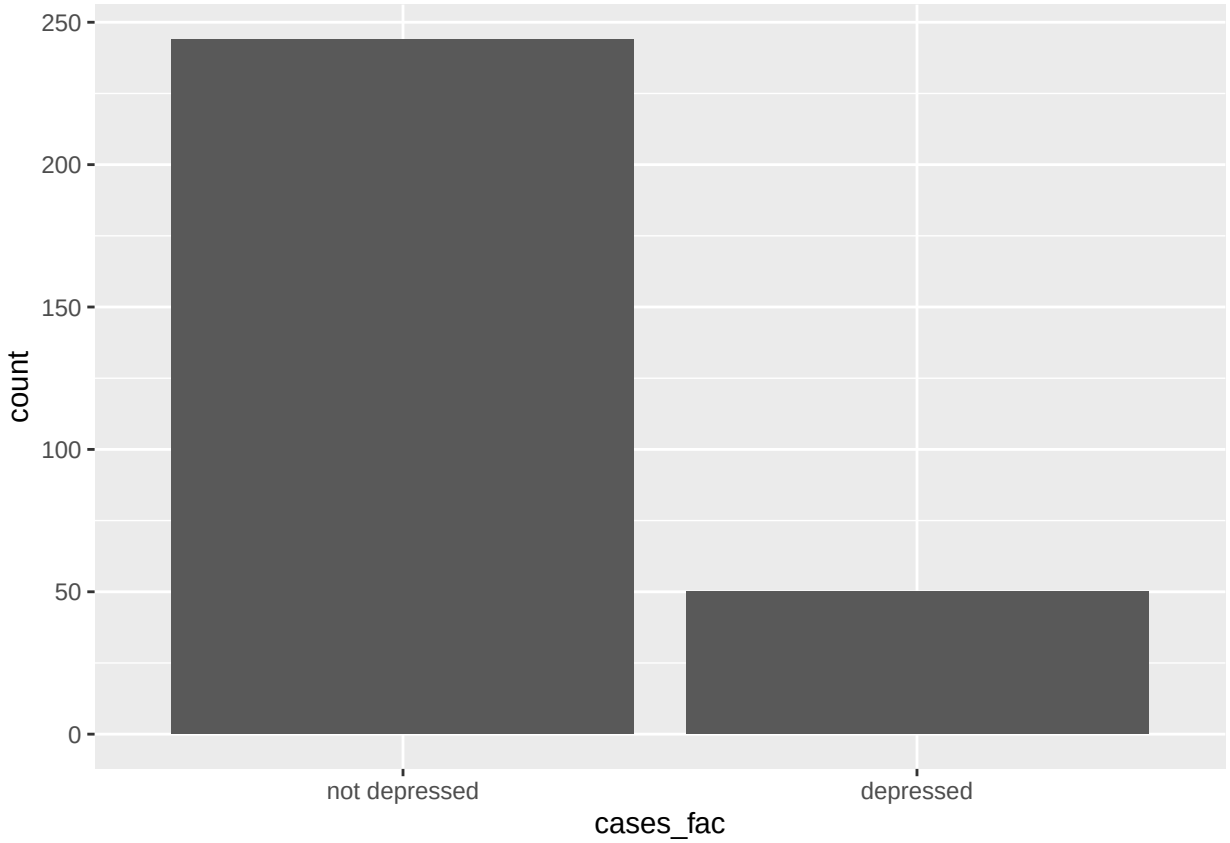
Depression Levels

```
depress$cases_fac <- factor(depress$CASES, labels=c("not depressed", "depressed"))
```

```
table(depress$cases_fac)
```

```
##
## not depressed     depressed
##           244            50
```

```
ggplot(depress, aes(x=cases_fac)) + geom_bar()
```



The majority of the individuals are not depressed. 244 individuals are not depressed and 50 of them do have depression.

In the above analysis of data you can see that in total there were 294 individuals surveyed. You can see that the dominant sex was female with 183 females total and only 111 males. In terms of employment status the majority were full-time employed with a total of 167 people being full-time employed, 42 being part-time employed, 14 being unemployed, 38 being retired, 27 being stay at home, 2 being students, and 4 were other not listed. In terms of marital status the dominant factor was married individuals with 127 people being married, 73 never been married, 43 divorced, 13 separated, and 38 widowed. In total there was 244 individuals who reported not being depressed and 50 being depressed so the majority were not depressed when surveyed.

Bivariate Comparison

SEX vs. CASES

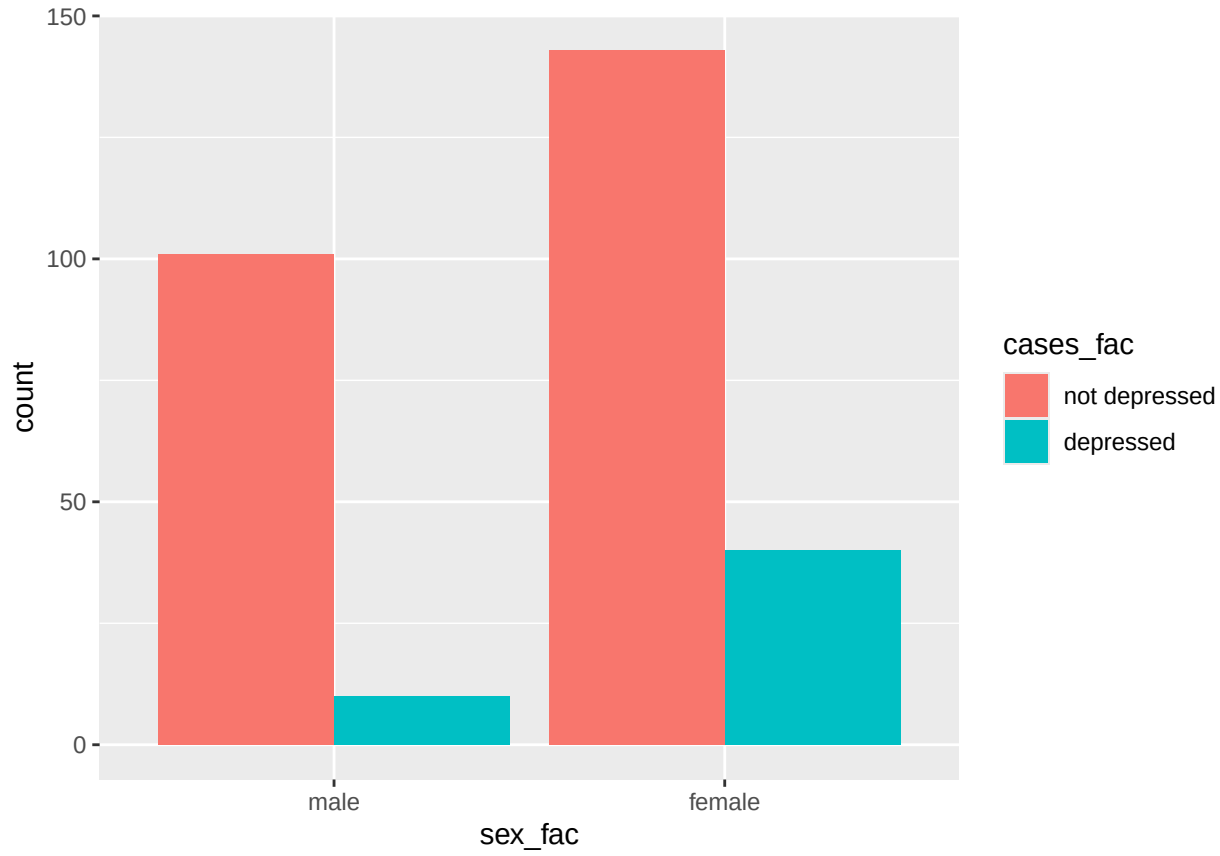
```
table(depress$sex_fac, depress$cases_fac)
```

```
##
##          not depressed depressed
##   male             101        10
##   female           143        40
```

```
prop.table(table(depress$sex_fac, depress$cases_fac), 1)
```

```
##
##          not depressed  depressed
##   male      0.90990991 0.09009009
##   female    0.78142077 0.21857923
```

```
ggplot(depress, aes(x=sex_fac, fill=cases_fac)) + geom_bar(position = "dodge")
```



40 women are depressed and only 10 men are depressed. Majority of them are not depressed.

MARITAL STATUS vs. CASES

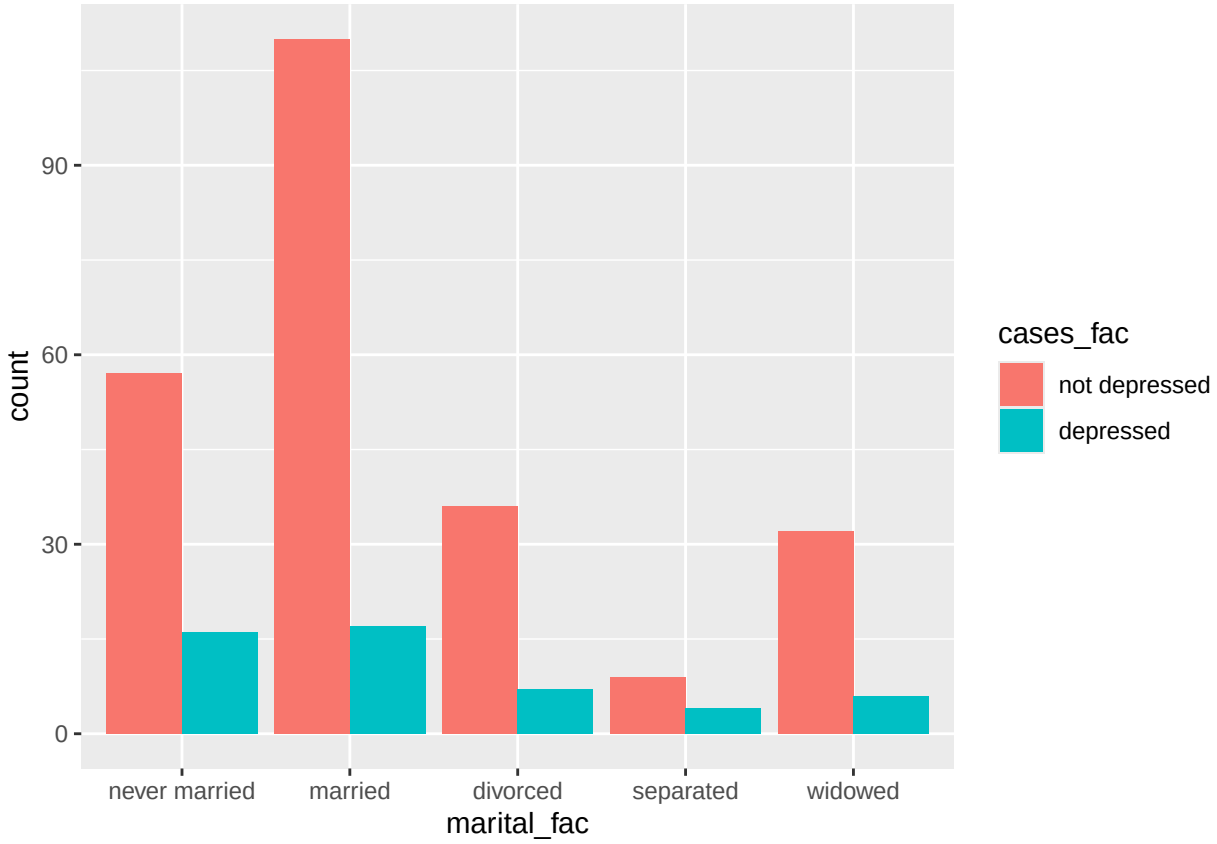
```
table(depress$marital_fac, depress$cases_fac)
```

```
##
##                 not depressed depressed
##   never married            57        16
##   married                 110        17
##   divorced                 36         7
##   separated                 9         4
##   widowed                  32         6
```

```
prop.table(table(depress$marital_fac, depress$cases_fac), 1)
```

```
##
##                 not depressed depressed
##   never married     0.7808219 0.2191781
##   married           0.8661417 0.1338583
##   divorced          0.8372093 0.1627907
##   separated         0.6923077 0.3076923
##   widowed           0.8421053 0.1578947
```

```
ggplot(depress, aes(x=marital_fac, fill=cases_fac)) + geom_bar(position = "dodge")
```



57 people who were not depressed have never been married, 110 of the people with no depression are married, 36 of the divorced don't have depression, 9 of the separated are not depressed, 32 of the widowed do not have depression. 16 of the people who have never been married are depressed, 17 of the people who are married have depression, 7 of the people who have been divorced have depression, and 6 of the people who have been widowed have depression.

EMPLOYMENT STATUS VS. CASES

```
table(depress$employ_fac, depress$cases_fac)
```

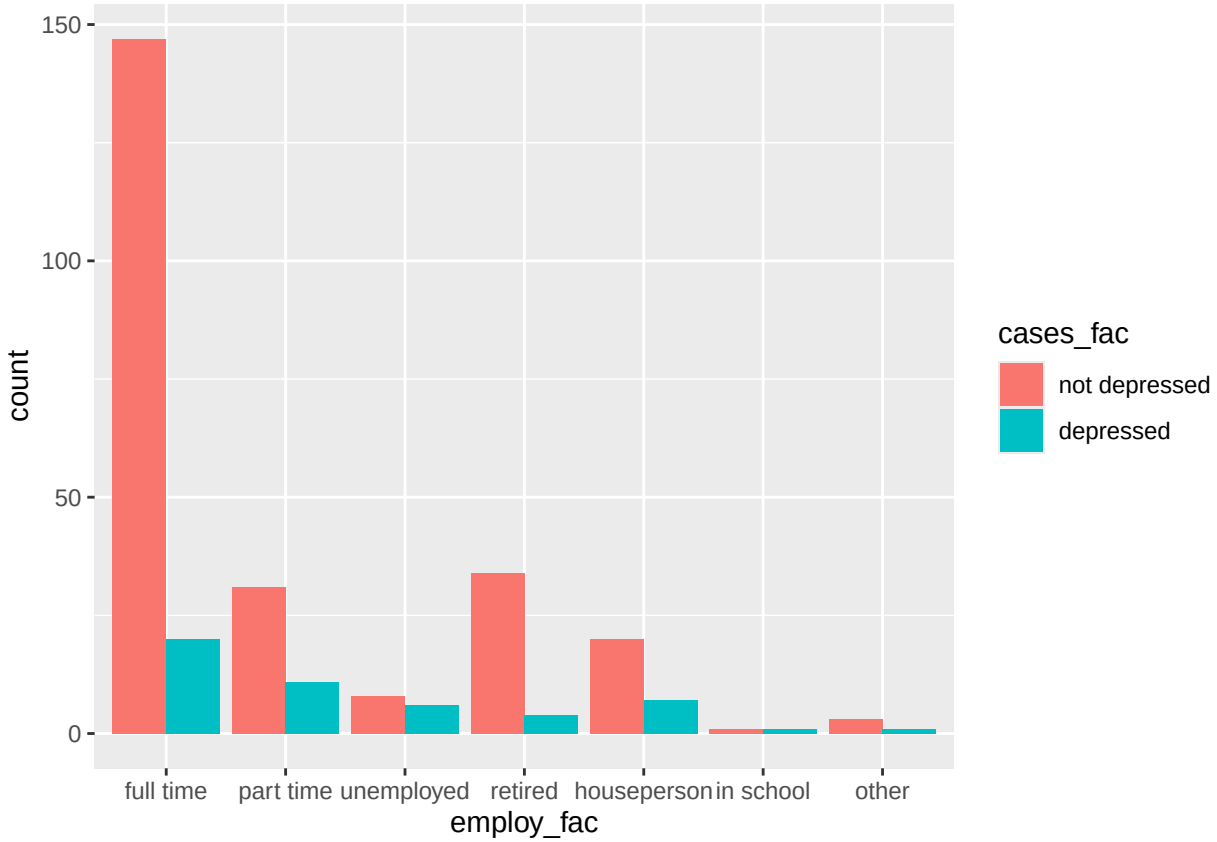
```
##
```

```
##                not depressed depressed
##   full time              147        20
##   part time               31        11
##   unemployed              8         6
##   retired                34         4
##   houseperson            20         7
##   in school               1         1
##   other                   3         1
```

```
prop.table(table(depress$employ_fac, depress$cases_fac), 1)
```

```
##
##                not depressed depressed
##   full time        0.8802395 0.1197605
##   part time        0.7380952 0.2619048
##   unemployed       0.5714286 0.4285714
##   retired          0.8947368 0.1052632
##   houseperson      0.7407407 0.2592593
##   in school        0.5000000 0.5000000
##   other            0.7500000 0.2500000
```

```
ggplot(depress, aes(x=employ_fac, fill=cases_fac)) + geom_bar(position = "dodge")
```



147 of the people who are full-time employees are not depressed. 31 of the people who are part-time employees are not depressed. 8 of the people who are unemployed are not depressed. 34 of the people who are retired are not depressed. 20 of the people who are housepeople do not have depression. 1 of the people

who are in school don't have depression. 3 of the people who are in the other category are not depressed. 20 of the people who are full-time employees are depressed, 11 of the part-time employees are depressed, 6 of the unemployed are depressed, 4 of the retired are depressed, 7 of the housepeople are depressed, 1 of the in school individuals are depressed, and 1 of the people in the other category are depressed.

Based on the data I observed that the majority of people in each category are not depressed. I can see that the highest numbers in depression cases in each category are in females with 40 of them being depressed, married people with a number of 17, and full-time employees with a number of 20.

Conclusion

After analyzing and exploring the data I found that only one of my hypothesis was correct which was that women will have higher cases of depression in comparison to men. The data does support this hypothesis because 40 females reported having depression in comparison to a lower number in men which was 10. My other two hypothesis were not supported by the data and proved otherwise. Overall I found that depression levels are lower than I thought. I had thought they would be higher since it is highly prevalent in society today. I found that the categories with higher numbers of depression are females, people who are married, and full-time employed individuals.